Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(1): 45-60.

Original Article

Multidimensional item response theory to assess psychometric properties of GHQ-12 in parents of school children

Elham Haem, Marziyeh Doostfatemeh*

Department of Biostatistics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran.

ARTICLE INFO

Received 01.08.2021
Revised 24.08.2021
Accepted 09.12.2021
Published 15.03.2022

Key words:
Multidimensional item response theory;
General health questionnaire;
Psychometrics properties;
Psychological distress.

ABSTRACT

Introduction: The multidimensional item response theory (MIRT) model provides an ideal foundation for assessing the psychological properties of a questionnaire designed with multidimensional structure. This study aimed to present the first use of MIRT models to investigate the psychometric properties of general health questionnaire (GHQ-12) in parents of school-aged children.

Methods: A total of 1104 parents of school children-aged completed the Persian version of GHQ-12 questionnaire. One-dimensional IRT model and MIRT models with two and three factors were used to model the observed scores for each of the GHQ-12 items as a function of the subject's latent characteristics to consider the correlation between the dimensions of the questionnaire. Goodness of fit indices were reported for the three models, and the fits of items were assessed for the best model. Individual items were described in detail through item characteristic curves, and the amount of information carried by different items was presented using information curves.

Results: The MIRT analysis with three factors corresponding with anxiety depression, social dysfunction and loss of confidence provided the best account of the GHQ-12 data. The model showed that all items were fitted adequately. Items varied in their discrimination ranged from 0.94 to 2.13, 1.31 to 2.74, and 2.87 to 3.57 for social dysfunction, anxiety depression, and loss of confidence, respectively. Moreover, items 8 and 2 provided the least information in social dysfunction and anxiety depression dimensions, respectively. Items in the loss of confidence dimension carried the most information among all items of the GHQ-12.

Conclusion: The developed framework for evaluating the psychometric properties of GHQ-12 can be a suitable alternative to traditional approaches as well as unidimensional IRT models, the use of which has been restricted due to the multidimensional structure of the questionnaire.

Introduction

The general health questionnaire (GHQ) is a self-report measure of minor psychiatric morbidity that has been widely used since its development by Goldberg in 1972.[1] The original instrument consists of 60 items, but different shorter versions, including GHQ-30,

*.Corresponding Author: doustma@sums.ac.ir

GHQ-28 and GHQ-12, have also been adapted and validated by different studies.[2] The 12-item version of the questionnaire, GHQ-12, has been broadly used because of its relatively good psychometric properties and brevity.[3, 4] Further, the GHQ-12 is recommended by the World Health Organization (WHO) as a well-validated and standard psychiatric screening instrument.[5]

The GHQ-12 consists of 12 items, each of which is rated on a four-point scale, typically worded: less than usual, no more than usual, rather more than usual, or much more than usual. The two most commonly used scoring methods are bi-modal (0-0-1-1) and Likert scoring styles (0-1-2-3).[6] Because the GHQ-12 exhibits considerable appeal as a quick and well-documented screening tool, it was translated into different languages to study its reliability and validity to explore its psychometric properties in various populations and countries.[6-12] For the first time, the Persian version of the questionnaire was prepared and its psychometric properties were assessed by Montazeri et al.[13] Since then, a couple of studies have assessed its applicability among university students and an elderly Iranian population.[5, 14]

The questionnaire was designed as a unidimensional scale to capture a single trait, and some empirical studies have supported this assumption.[15, 16] However, studies have frequently revealed the existence of more than one factor solutions. Most studies have yielded a two-factor solution, namely "anxiety/depression" and "social dysfunction".[7, 17-22]

Some studies, however, have revealed a third factor expressing "loss of confidence".[23-26] For the Persian version of the questionnaire, a two-factor model was the best explanation of the Iranian sample.[13]

Traditionally, classical test theory (CTT) measuring construct validity, reproducibility, and sensitivity to change, is used to assess the psychometric properties of questionnaires.[27] Furthermore, confirmatory factor analysis (CFA), is a common method, used to evaluate hypotheses about the dimensionality of a questionnaire.[28] Typically, these techniques are used to evaluate a new questionnaire in a first step. Although these methods are beneficial in assessing the psychometric properties and scale-dimensionality of a new questionnaire, more modern methods based on item response theory (IRT) framework exist that provide a more detailed assessment of individual items in a questionnaire.

IRT, also known as the latent response theory attempts to explain the relationship between an individual response to the items on a questionnaire and the latent trait.[29, 30] It establishes a link between the properties of items on a questionnaire, individuals responding to these items, and the underlying trait being measured.

Despite IRT benefits, most studies on the psychometric properties of GHQ-12 have used CTT methods, exploratory factor analysis (EFA) and CFA. However, a couple studies used a unidimensional IRT model to assess the hypothesis on the factorial structure of GHQ-12.[31-33] Alexandrowicz et al.[34] applied unidimensional IRT models with the aim of comparing the 30-, 20-, and 12-item versions of GHQ with four different recording schemes. When questionnaires comprise multiple dimensions, the utility of unidimensional IRT is largely restricted. An improved version of the IRT model, multidimensional IRT (MIRT), takes multiple latent traits into account simultaneously while also considering correlation among latent traits. MIRT models have rarely been used in GHQ-12, although

such studies had different aims.[35, 36]

It appears that there is no reported MIRT-based study on the psychiatric morbidity of the parents of school-aged children measured by GHQ-12. Whereas children's quality of life is one of the important and complementary outcomes in clinical studies, several studies have focused on this subject.[37-39] The health-related quality of life in children is strongly influenced by the mental health of their parents. Therefore, it is also crucial to evaluate the parents' psychiatric morbidity in a population of children.

The present study aimed to use MIRT models to investigate the properties of the questionnaire with more detail. The unidimensional IRT and MIRT models with two and three factors were applied to the data and the three models were compared to each other using several goodness of fit indices. Afterwards, in the best-fitted model, individual items were described in detail through item characteristic curves and item information curves.

Material and Methods

Participants and instrument

The Persian version of the GHQ-12, previously translated and validated in Iran,[13] was completed by 1104 parents of Iranian secondary school adolescents aged 13-18 years. A two-stage cluster random sampling technique was used to select the participants randomly. In the first stage, four schools were randomly selected from 60 secondary schools in each of the four educational districts in Shiraz, southern Iran. In the second stage, two classes from each school were chosen through a simple random sampling and all parents of the students in the chosen classes were considered as the study population. The students took the informed consent forms and the questionnaires home for their parents, and then the completed questionnaires were returned to the schools.

The GHQ-12 includes 12 ordered categorical questions or items which are rated in four categories 0, 1, 2 and 3, indicating less than usual, no more than usual, rather more than usual, or much more than usual, respectively. The GHQ-12 scoring protocol has reversed-scored items such that higher scores show a better state of psychological health, and the model was fitted accordingly.

Multidimensional item response theory

IRT models assume that there is only one latent variable, θ, to explain the relationship between latent traits and observed responses. However, MIRT, as an extension of IRT models, attempts to explain an item response according to an individual's standing multiple latent dimensions.[40] Several forms of IRT models have been used for ordered categorical data: the rating scale model, partial credit model, generalized partial credit model (GPCM), and graded response model (GRM).[30] The most common IRT-based approach for multiple-response questionnaires in patient-reported outcome studies has been GRM.[29]

The functional form of the multidimensional GRM is given by:

$$P(Y_{ij} \geq K \mid \theta_i = \theta) = \frac{1}{1+exp\left[-\left(a_j^T \theta + c_{jk}\right)\right]} \quad (1)$$

$$P(Y_{ij} = K \mid \theta_i = \theta) = P(Y_{ij} \geq K \mid \theta_i = \theta) - P(Y_{ij} \geq K+1 \mid \theta_i = \theta) \quad (2)$$

where $P(Y_{ij} \geq K \mid \theta_i = \theta)$ is the probability that observed scores for item j and subject i given the ability on latent trait θ_i obtain a score greater or equal to k, with k=0 to 3. In this equation, a_j and

c_{jk} denote the item discrimination and intercept, respectively, where intercepts are ordered and one less than the number of response categories for each item. A high discrimination value shows that an item is able to differentiate between the subjects at different latent trait levels. The intercept, c_{jk}, can be transformed into a difficulty parameter, b_{jk}, through the following formula:

$$b_{jk} = \frac{-c_{jk}}{a_j} \quad (3)$$

where a low value for the difficulty parameter indicates an easy item and a high value of difficulty indicates a difficult item. Furthermore, in Eq,[1] latent traits are distributed normally, $\theta_i \sim N(0,\Omega)$, where Ω is the covariance matrix for individual i's latent traits. The correlation between the dimensions is taken into account in the multidimensional GRM model through Ω.[28, 41]

It was said that, the GHQ-12 was originally designed as a unidimensional scale. Thereafter, several studies explored two-[7, 17-22] and even three-factor[23-25] solutions. The reliability and validity of the Persian version of the GHQ-12 was evaluated by Montazeri et al.,[13] and a two-factor model was revealed as the best explanation of the questionnaire. Thus, according to the mentioned study, an MIRT model including two factors, namely social dysfunction (items 1, 3, 4, 7, 8, 10, and 11) and psychological distress (items 2, 5, 6, 9, and 12) was applied. Because, a couple of studies reported a three-factor solution as the best model, an exploratory factor analysis (EFA) was performed to manifest the best factor structure on the GHQ-12. The EFA showed a three-factor structure as the best explanation of the data. Afterwards, an MIRT model with three factors namely social dysfunction (including items 1, 3, 4, 7, 8, and 12), anxiety depression (including items 2, 5, 6, and 9), and loss of confidence (including items 10 and 11), was performed. Furthermore, a unidimensional IRT model was applied and compared with MIRT models that have two and three factors.

Statistical analysis

All analyses were performed in the R programming environment. The EFA and MIRT models were performed using *stats*[42] and multidimensional item response theory (*mirt*) packages,[43] respectively. The unidimensional IRT model and the MIRT models with two and three factors were compared using Akaike information criterion (AIC). Furthermore, the goodness of fit of the models was evaluated by the comparative fit index (CFI), Tucker- Lewis Index (TLI), and the root-mean-square error of approximation (RMSEA). The following cut-off values for a good fit were suggested by Hu and Bentler[44]: CFI>0.95,TLI>0.95,RMSEA<0.06.

Item characteristic curves (ICC) were provided to describe the probability of each score in each item visually. Furthermore, item information curves were included to investigate which items of GHQ-12 carried the most information to detect psychiatric morbidity in the parents. Information content was calculated for the items using Fisher information, which is formulated minus the expectation of the second derivative of the log-likelihood of the model.[29] To evaluate the item fit, the generalized Orlando and Thissen's S-X^2 index for polytomous data was used,[45] comparing the observed and expected response frequencies under the estimated MIRT model. Eventually, the items with S-X^2 p-value<0.01 were considered poorly fitted.[46, 47]

Results

In this study, there were 13,248 observations from 1104 parents of school-aged children. The mean± standard deviation of the parents' age was 42.5± 6.2, and 57% of the parents had academic degrees.

Unidimensional IRT and MIRT models with two and three factors were fitted on the GHQ-12 data set. Table 1 summarizes the goodness of fit of the models, representing the MIRT model with three factors namely anxiety depression, social dysfunction, and loss of confidence, which reflected the data better compared to the other models. This model had the lowest AIC and met cut off values for a good fit. Thus, the MIRT model with three factors was considered for further evaluation.

The distributions of the observed responses of items for the three dimensions are shown in Figure 1. The frequency of ordinal items showed diverse patterns in the three dimensions. In the social dysfunction and loss of confidence dimensions, most items were skewed toward high scores (2 or more), indicating a better state of psychological health, while items of anxiety depression were more symmetrically distributed.

In the MIRT model with three factors, item specific parameters and the correlation between the three factors were estimated successfully. Table 2 displays the estimation of item discrimination, item difficulty parameters and correlations among the three factors. For all items in the three dimensions, discrimination estimates ranged from 0.94 to 3.57, indicating that apart from item 8, all items discriminated between low and high levels of GHQ-12 latent traits (or state of psychological health) of the parents very well. Furthermore, the estimated correlations among the three factors ranged from 0.70 to 0.76, showing that an increase in one of the latent traits leads to an increase in others.

Figure 2 shows the obtained ICCs for all items in the GHQ-12. This figure indicates that a person with a better state of psychological health (higher latent trait of anxiety depression, social dysfunction, or loss of confidence) has a higher probability of increased scores for each item. The lowest slope of 0.94 for face up to problems (item 8) indicates the lower discrimination power in social dysfunction of parents. In other words, a large increment in health state yields only a small increment in the probability for the score on this item. However, the high slope parameter of 2.87 and 3.57 for thinking of self as worthless (item11) and losing confidence (item 10) indicates a higher discrimination power in loss of confidence trait. For all items, when the psychological health state score increases, the probability of a 0 score decreases.

Figure 3 presents the item information curves for all items of the anxiety depression, social dysfunction, and loss of confidence dimensions

Table 1. MIRT model fit

Fit statistics	One-dimensional	Two-dimensional	Three- dimensional
Log-likelihood	-12623.96	-12589.71	-12464.23
AIC	25343.92	25277.41	25030.44
RMSEA (CI.90%)	0.04 (0.03, 0.05)	0.04 (0.03-0.05)	0.05(0.04, 0.06)
CFI	0.96	0.96	0.96
TLI	0.94	0.95	0.95

AIC, Akaike information criterion; RMSEA,Root mean square error of approximation; CI, Confidence interval; CFI, Comparative fit index; TLI, Tucker- Lewis index

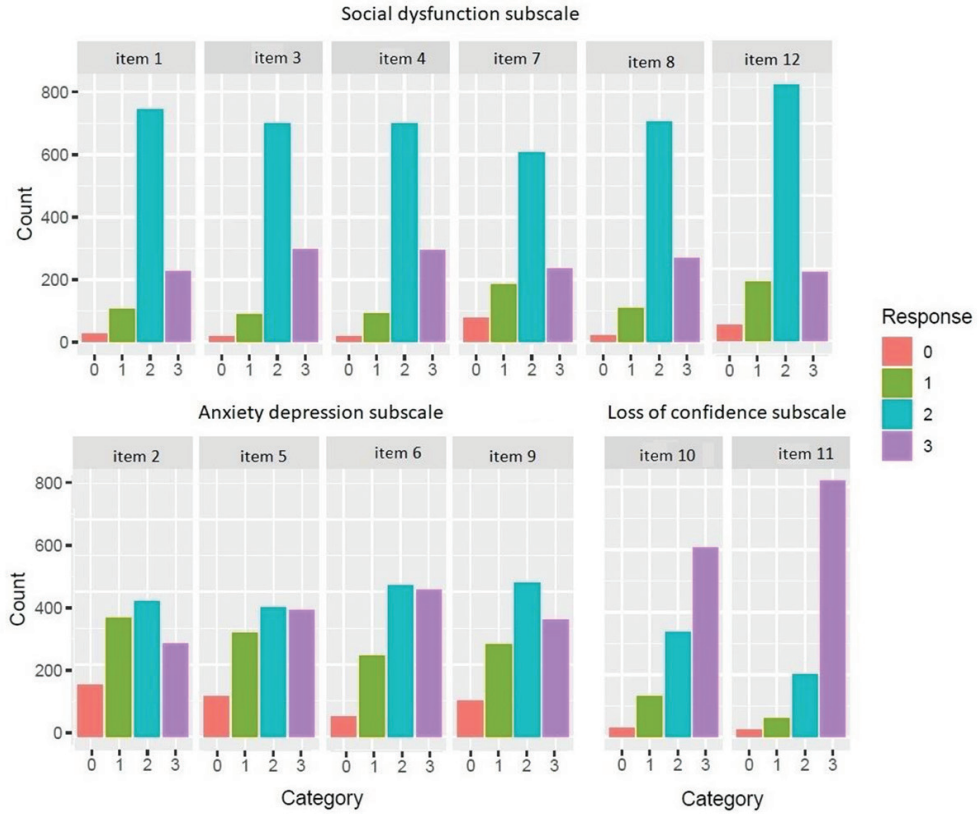

Figure 1. Distributions of observed item responses (0= much more than usual, 1= rather more than usual, 2= no more than usual, 3= less than usual) for each dimension. The names of the item are provided in Table 2.

Table 2. MIRT with results presented as estimate

MIRT parameter estimates for the graded response model						
item	Discrimination			Difficulty		
	a_1	a_2	a_3	b_1	b_2	b_3
1. Able to concentrate	-	1.45	-	-3.27	-1.84	1.25
2. Lost much sleep	1.31	-	-	-1.88	-0.32	1.15
3. Playing useful part	-	1.37	-	-3.65	-2.10	0.97
4. Capable of making decisions	-	1.22	-	-3.94	-2.23	1.057
5. Under stress	1.62	0	-	-1.88	-0.51	0.67
6. Could not overcome difficulties	1.57	0	-	-2.48	-0.98	0.47
7. Enjoy normal activities	-	1.89	-	-2.07	-0.97	1.04
8. Face up to problems	-	0.94	-	-4.61	-2.45	1.41
9. Feeling unhappy and depressed	2.74	-	-	-1.59	-0.53	0.61
10. Losing confidence	-	-	3.57	-2.24	-1.17	-0.16
11. Thinking of self as worthless	-	-	2.87	-2.47	-1.68	-0.77
12. Feeling reasonably happy	-	2.13	-	-2.34	-1.13	1.19
Correlations for psychological distress, social dysfunction			0.71			
social dysfunction, loss of confidence			0.70			
psychological distress, loss of confidence			0.76			

a_1 is the discrimination parameter of the corresponding item on anxiety depression dimension, a_2 is the discrimination parameter of the corresponding item on social dysfunction dimension and a_3 is loss of confidence dimension

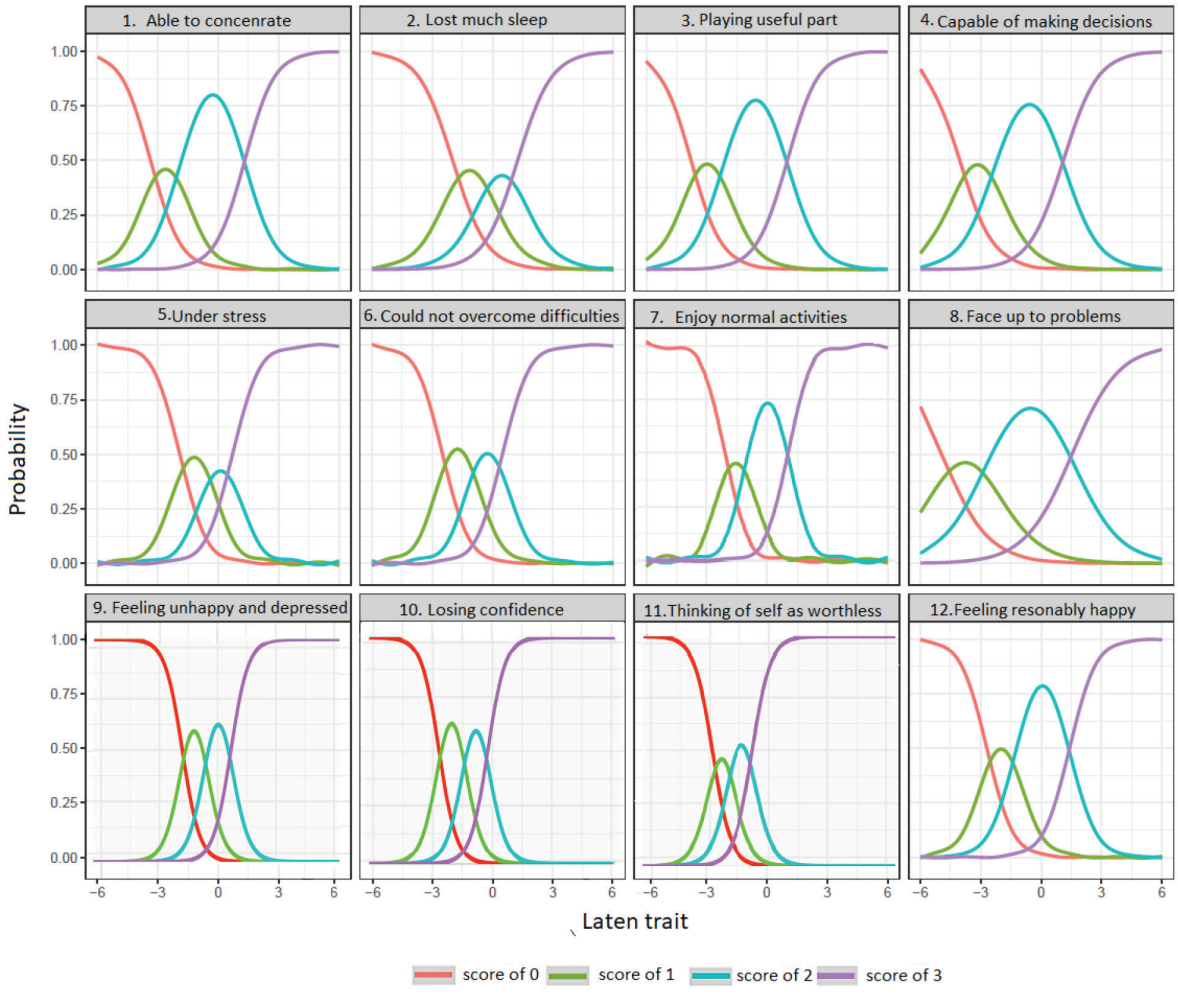

Figure 2. Item characteristic curves showing the probability for each individual score within each category of items.

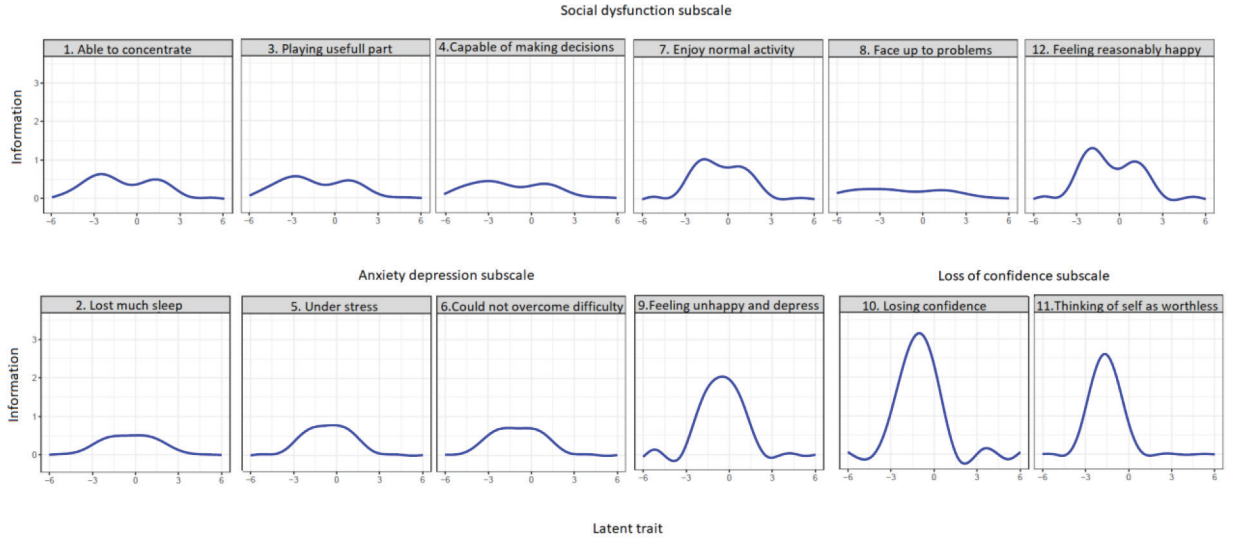

Figure 3. Item information curves for items of anxiety depression, social dysfunction and loss of confidence dimensions.

Table 3. Item fit statistics

Item	S-X^2 [a]	df	p-value[b]
1. Able to concentrate	35.64	39	0.62
2. Lost much sleep	52.92	51	0.44
3. Playing useful part	49.41	34	0.08
4. Capable of making decisions	57.53	33	0.03
5. Under stress	66.18	47	0.08
6. Could not overcome difficulties	56.70	47	0.21
7. Enjoy normal activities	51.71	41	0.21
8. Face up to problems	46.74	39	0.22
9. Feeling unhappy and depressed	51.45	34	0.08
10. Losing confidence	42.59	34	0.21
11. Thinking of self as worthless	59.99	34	0.03
12. Feeling reasonably happy	59.34	39	0.08

[a] Orlando and Thissen's S-X^2 index, items with S-X^2 p-value<0.01 are considered poorly fitted.
[b] p-values were rounded in two digits.

separately. Item information curves, as a function of latent traits, indicate which item carries the most information and where on the latent trait they are most informative. The information content carried by items differed. In social dysfunction, feeling reasonably happy (item 12) and enjoy normal activities (item 7) were the most informative over the moderate range of latent trait, while face up to problems (item 8) carried little to almost no information in this study. Moreover, in anxiety depression, feeling unhappy and depressed (item 9) carried the most information on the moderate latent trait. However, lost much sleep (item 2) was the least informative over a broad range of the latent trait. Furthermore, the loss of confidence dimension, including losing confidence (item 10) and thinking of self as worthless (item 11), were the most informative among all items, respectively.

Table 3 shows the full results for item fit statistics. Based on S-X^2 p-values, all the items fit the GHQ-12 questionnaire properly.

Discussion

The present study is the first to apply the MIRT model to evaluate the psychometric properties of the GHQ-12 questionnaire in parents of school-aged children. This study included 1104 parents to measure their minor psychiatric morbidity. Since both maternal and paternal psychological health affects the children's development and health during school, assessment of the parents' psychiatric morbidity is essential.

The analysis of questionnaires and the assessment of their psychometric properties using the CTT approach which focuses on summated scores disregards the underlying nature of the data. EFA and CFA procedures are predominantly used to assess the dimensionality or underlying latent variable structure of a questionnaire. However, MIRT models provide more detail on a questionnaire and yield potentially rich and informative results about individual items. Once a model is selected, ICC curves visually present the power of discrimination and difficulty of individual items in subscales. Item information functions are obtained through IRT models and estimate the precision and reliability of individual items independent of other items on the questionnaire. In addition, item information curves indicate the content of information carried by individual

items. As a result, a subset of items can be selected, and a reduced questionnaire can be developed by omitting uninformative items.

Notwithstanding the benefits of IRT, it suffers from one limitation which is the need for large samples. A summary of the recommended sample sizes for various IRT models has been provided by Yen and Fitzpatrick.[48] MIRT, as an extension of the IRT approach, models multiple dimensions simultaneously to take the correlations among the dimensions into account. Because these correlation parameters are estimated among the dimensions, MIRT models need a larger sample compared to IRT models. In this study, a sufficiently large sample was employed to obtain stable parameter estimates in the MIRT model.

In the present study, the MIRT model with three factors reflected the data better than the other models. The current findings were in line with other studies that have reported three-dimensional structure, including anxiety depression, social dysfunction, and loss of confidence, although they used CTT and CFA.[23, 25,26] Smith et al.[31] applied a Rasch model and CFA to the 12-item GHQ and identified 6 misfitting items. They focused more on differential item functioning by age, gender, and treatment aims. However, the discrimination and difficulty parameters, ICC and information curves were not reported.[31] The current findings highlight no misfitting items, and this result is not in line with the mentioned study. This inconsistency may be explained by the differences between MIRT models, considering correlations among dimensions, and unidimensional IRT models. Furthermore, the current study, employed a graded response model through the MIRT model, while Smith et al.[31] applied the Rasch model in the IRT approach. Since graded response models have fewer assumptions compared to Rasch models, they are more flexible and likely to fit the data generated from the patients' reported outcomes.[49]

As it was note before, MIRT models are seldom applied on GHQ-12. Stochl et al.[35] combined GHQ-12 and Affectometer-2 in an item bank through the computerized adaptive testing method for public mental health research. They applied the MIRT model on pooled items and reported that the proposed item bank was more efficient than the use of either measure alone. The current findings are not comparable with the mentioned study, because the MIRT model was not applied in the two questionnaires separately. In another study, items of three instruments, namely the Warwick-Edinburgh Mental Well-being Scale (WEMWBS), EQ-5D items (Health Survey for England) and GHQ-12, were combined, and unidimensional IRT and MIRT models with two and three factors were performed.[36] That study aimed to establish the number of latent variables needed to explain the responses to all of the items on the three questionnaires and also to investigate relationship between latent variables. Thus, the results of the mentioned study should not be compared with the current findings. To the best of the authors' knowledge, the present study is the first to apply MIRT models only on GHQ-12 items and to provide details on individual items through ICC and information curves.

As mentioned before, an advantage of IRT-based models is the amount of item information calculated based on item characteristic curves. This provides the relative contribution of different items to total information across different regions along the latent trait. Consequently, item information curves play a significant role in the description of items, optimal selection of the most informative subset of items, and in comparing efficiency

rates between different tests.[29, 50] In the social dysfunction dimension, two items, i.e face up to problems (item 8) and capable of making decisions (item 4) were found to have the least information. In the anxiety depression dimension, lost much sleep (item 2) included lower information in a broad range of the latent trait compared to other items. In the loss of confidence dimension, both items carried the most information among the 12 items of the questionnaire. Hence, a subset of more informative items can be selected, and a shortened version of GHQ-12 can be developed.

The present study had a number of limitations which should be taken into consideration. First, the participants were from a general population. Thus, the results cannot be extended to subgroups suffering serious chronic illnesses. Second, the participants in this study consisted of fathers and mothers of school-aged children. Fathers and mothers probably have different perceptions of specific items in the GHQ-12 questionnaire; Thus, methodologically, combining them may be misleading.[37, 51] Therefore, the measurement invariance of the GHQ-12 across fathers and mothers should be assessed in future studies. The third limitation of this study was that the estimation of the MIRT parameters was not adjusted according to cluster sampling. However, in this study, the number of cluster participants was almost the same in each cluster and a simulation study by Lee et al.,[52] indicated that the two-stage cluster estimator should be used when the number of participants per cluster is significantly different. Finally, it is recommended that future studies should address these limitations and try to expand the current findings in the GHQ-12 to different subgroups.

Conclusion

Based on GHQ-12 data from the parents of school-aged children, an MIRT model with three factors, namely anxiety depression, social dysfunction, and loss of confidence, was successfully developed to examine the psychometric properties of the questionnaire. Additionally, item fit statistics assessed individual items, and information curves described the amount of information carried by individual items. MIRT models can be adapted as powerful tools to examine the psychometric properties of questionnaires designed with an intentional multidimensional structure. It is hoped that the published articles on MIRT models stimulate its increased use in the field of health psychology.

Declarations

Ethics approval and consent to participate

This study was approved by the local ethics committee of Shiraz University of Medical Sciences. We confirmed that all methods were performed in accordance with the relevant guidelines and regulations. All participants were informed about the study and confidentiality protocols. Written informed consent was taken from all the participants.

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study is available from the corresponding author on reasonable request.

Conflicts of interest

The authors declare that they have no competing interests.

Funding

This work was supported by no funding

Authors' contributions

Elham Haem analyzed and wrote the manuscript and researched the data, and Marziyeh Doostfatemeh analyzed and researched the data. Both authors read and approved the final manuscript.

Acknowledgment

The authors wish to thank the Radan-English-Edit Institute for the invaluable assistance in language editing of this manuscript.

References

1. Goldberg DP, Blackwell B. Psychiatric illness in general practice: a detailed study using a new method of case identification. Br med J. 1970;2(5707):439-43.

2. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychological Medicine. 2009;9(1):139-45. Epub 07/09.

3. Romppel M, Braehler E, Roth M, Glaesmer H. What is the General Health Questionnaire-12 assessing?: Dimensionality and psychometric properties of the General Health Questionnaire-12 in a large scale German population sample. Comprehensive Psychiatry. 2013 2013/05/01/;54(4):406-13.

4. Salama-Younes M, Montazeri A, Ismaïl A, Roncin C. Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: a study from France. Health Qual Life Outcomes [Internet]. 2009 2009; 7:[22 p.]. Available from: http://europepmc.org/abstract/MED/19265516
http://www.biomedcentral.com/1477-7525/7/22/abstract
http://www.biomedcentral.com/content/pdf/1477-7525-7-22.pdf
http://www.biomedcentral.com/1477-7525/7/22
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/19265516/?tool=EBI
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/19265516/pdf/?tool=EBI
https://doi.org/10.1186/1477-7525-7-22
https://europepmc.org/articles/PMC2653033
https://europepmc.org/articles/PMC2653033?pdf=render.

5. Rahmati Najarkolaei F, Raiisi F, Rahnama P, Gholami Fesharaki M, Zamani O, Jafari MR, et al. Factor structure of the Iranian version of 12-item general health questionnaire. Iran Red Crescent Med J. 2014;16(9):e11794-e. PubMed PMID: 25593708. eng.

6. Liang Y, Wang L, Yin X. The factor structure of the 12-item general health questionnaire (GHQ-12) in young Chinese civil servants. Health Qual Life Outcomes. 2016;14(1):136-. PubMed PMID: 27669741. eng.

7. Politi PL, Piccinelli M, Wilkinson G. Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. Acta Psychiatrica Scandinavica. 1994 1994/12/01;90(6):432-7.

8. Daradkeh TK, Ghubash R, el-Rufaie OE. Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. Psychol Rep. 2001 2001/08//;89(1):85-94. PubMed PMID: 11729557. eng.

9. Quek KF, Low WY, Razack AH, Loh CS. Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: a Malaysian study. Psychiatry Clin Neurosci. 2001 2001/10//;55(5):509-13. PubMed PMID: 11555347. eng.

10. Kilic C, Rezaki M, Rezaki B, Kaplan I, Ozgen G, Sağduyu A, et al. General Health Questionnaire (GHQ12 & GHQ28): psychometric properties and factor structure of the scales in a Turkish primary care sample. Soc Psychiatry Psychiatr Epidemiol. 1997 1997/08//;32(6):327-31. PubMed PMID: 9299926. eng.

11. Chan DW. The Chinese General Health Questionnaire in a psychiatric setting: the development of the Chinese scaled version. Soc Psychiatry Psychiatr Epidemiol. 1993 1993/07//;28(3):124-9. PubMed PMID: 8378807. eng.

12. Sánchez-López MdP, Dresch V. The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. Psicothema. 2008 2008/11//;20(4):839-43. PubMed PMID: 18940092. eng.

13. Montazeri A, Harirchi AM, Shariati M, Garmaroudi G, Ebadi M, Fateh A. The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. Health Qual Life Outcomes. 2003;1:66-. PubMed PMID: 14614778. eng.

14. Namjoo S, Shaghaghi A, Sarbaksh P, Allahverdipour H, Pakpour AH. Psychometric properties of the General Health Questionnaire (GHQ-12) to be applied for the Iranian elder population. Aging & Mental Health. 2017 2017/10/03;21(10):1047-51.

15. Fernandes HM, Vasconcelos-Raposo J. Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. Assessment. 2013 2013/04//;20(2):219-29. PubMed PMID: 23118267. eng.

16. Hahn D, Reuter K, Härter M. Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. Psychosoc Med [Internet]. 2006 2006; 3:[Doc09 p.]. Available from: http://europepmc.org/abstract/MED/19742274 https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/19742274/?tool=EBI https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/19742274/pdf/?tool=EBI https://europepmc.org/articles/PMC2736509 https://europepmc.org/articles/PMC2736509?pdf=render.

17. Vanheule S, Bogaerts S. The factorial structure of the GHQ-12. Stress and Health: Journal of the International Society for the Investigation of Stress. 2005;21(4):217-22.

18. Toyabe S-i, Shioiri T, Kobayashi K, Kuwabara H, Koizumi M, Endo T, et al. Factor structure of the General Health Questionnaire (GHQ-12) in subjects who had suffered from the 2004 Niigata-Chuetsu Earthquake in Japan: a community-based study. BMC Public Health. 2007;7(1):175.

19. Schmitz N, Kruse J, Tress W. Improving screening for mental disorders in the primary care setting by combining the GHQ-12 and SCL-90-R subscales. Comprehensive psychiatry. 2001 2001 Mar-Apr;42(2):166-73. PubMed PMID: 11244154. eng.

20. Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ-12. J Eur Acad Dermatol Venereol. 2001 2001/09//;15(5):410-7. PubMed PMID: 11763380. eng.

21. Kalliath TJ, O'Driscoll MP, Brough P. A confirmatory factor analysis of the General Health Questionnaire-12. Stress and Health. 2004 2004/02/01;20(1):11-20.

22. Gureje O. Reliability and the factor structure of the Yoruba version of the 12-item General Health Questionnaire. Acta Psychiatrica Scandinavica. 1991 1991/08/01;84(2):125-9.

23. Campbell A, Walker J, Farrell G. Confirmatory factor analysis of the GHQ-12: can I see that again? Aust N Z J Psychiatry. 2003 2003/08//;37(4):475-83. PubMed PMID: 12873334. eng.

24. French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. Eur Child Adolesc Psychiatry. 2004 2004/02//;13(1):1-7. PubMed PMID: 14991426. eng.

25. Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. Psychol Assess. 2005 2005/06//;17(2):231-6. PubMed PMID: 16029110. eng.

26. Graetz B. Multidimensional properties of the General Health Questionnaire. Soc Psychiatry Psychiatr Epidemiol. 1991 1991/05//;26(3):132-8. PubMed PMID: 1887291. eng.

27. Goetz C, Ecosse E, Rat A-C, Pouchot J, Coste J, Guillemin F. Measurement properties of the osteoarthritis of knee and hip quality of life OAKHQOL questionnaire: an item response theory analysis. Rheumatology (Oxford). 2011 2011/03//;50(3):500-5. PubMed PMID: 21071480. eng.

28. Depaoli S, Tiemensma J, Felt JM. Assessment of health surveys: fitting a multidimensional graded response model. Psychology, Health & Medicine. 2018 2018/12/14;23(sup1):1299-317.

29. Haem E, Doostfatemeh M, Firouzabadi N, Ghazanfari N, Karlsson MO. A longitudinal item response model for Aberrant Behavior Checklist (ABC) data from children with autism. J Pharmacokinet Pharmacodyn. 2020 2020/04//. PubMed PMID: 32285302. eng.

30. Doostfatemeh M, Taghi Ayatollah SM, Jafari P, Jafari P. Power and Sample Size Calculations in Clinical Trials with Patient-

Reported Outcomes under Equal and Unequal Group Sizes Based on Graded Response Model: A Simulation Study. Value Health. 2016 2016 Jul-Aug;19(5):639-47. PubMed PMID: 27565281. eng.

31. Smith AB, Fallowfield LJ, Fallowfield LJ, Stark DP, Velikova G, Jenkins V. A Rasch and confirmatory factor analysis of the general health questionnaire (GHQ)--12. Health Qual Life Outcomes [Internet]. 2010 2010/04//; 8:[45 p.]. Available from: http://europepmc.org/abstract/MED/20433690
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/20433690/?tool=EBI
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/20433690/pdf/?tool=EBI
http://www.biomedcentral.com/content/pdf/1477-7525-8-45.pdf
http://www.biomedcentral.com/1477-7525/8/45
http://www.biomedcentral.com/1477-7525/8/45/abstract
https://doi.org/10.1186/1477-7525-8-45
https://europepmc.org/articles/PMC2873594
https://europepmc.org/articles/PMC2873594?pdf=render.

32. Doyle F, Watson R, Morgan K, McBride O. A hierarchy of distress and invariant item ordering in the General Health Questionnaire-12. Journal of Affective Disorders. 2012 2012/06/01/;139(1):85-8.

33. Mayhew E, Stuttard L, Beresford B. An Assessment of the Psychometric Properties of the GHQ-12 in an English Population of Autistic Adults Without Learning Difficulties. J Autism Dev Disord. 2020 2020/07//. PubMed PMID: 32653976. eng.

34. Alexandrowicz RW, Friedrich F, Jahn R, Soulier N. Using Rasch-models to compare the 30-, 20-, and 12-items version of the general health questionnaire taking four recoding schemes into account. Neuropsychiatr. 2015 2015;29(4):179-91. PubMed PMID: 26512001. eng.

35. Stochl J, Böhnke JR, Pickett KE, Croudace TJ. An evaluation of computerized adaptive testing for general psychological distress: combining GHQ-12 and Affectometer-2 in an item bank for public mental health research. BMC Med Res Methodol [Internet]. 2016 2016/05//; 16:[58 p.]. Available from: http://europepmc.org/abstract/MED/27206714
https://doi.org/10.1186/s12874-016-0158-7
https://europepmc.org/articles/PMC4875708
https://europepmc.org/articles/PMC4875708?pdf=render.

36. Böhnke JR, Croudace TJ. Calibrating well-being, quality of life and common mental disorder items: psychometric epidemiology in public mental health research. Br J Psychiatry. 2016 2016/08//;209(2):162-8. PubMed PMID: 26635327. eng.

37. Doostfatemeh M, Ayatollahi SMT, Jafari P, Jafari P. Testing parent dyad interchangeability in the parent proxy-report of PedsQL™ 4.0: a differential item functioning analysis. Qual Life Res. 2015 2015/08//;24(8):1939-47. PubMed PMID: 25645741. eng.

38. Christine E, Morse R. Can Parents Rate Their Child's Health-Related Quality of Life? Results of a Systematic Review. Quality of Life Research. 2001;10(4):347-57.

39. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes. 2007;5:2-. PubMed PMID: 17201923. eng.

40. te Marvelde JM, Glas CA, Van Landeghem G, Van Damme J. Application of multidimensional item response theory models to longitudinal data. Educational and Psychological Measurement. 2006;66(1):5-34.

41. Doostfatemeh M, Haem E, Sarbaraninan M, Ajdari Tafti M. Multidimensional item Response theory to assess the psychometric properties of persian version of dutch eating behavior questionnaire (DEBQ) in university students. Current Psychology. 2022:1-11.

42. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

43. Chalmers RP. mirt: A Multidimensional Item Response Theory Package for the R Environment. 2012. 2012 2012-05-24;48(6):29. Epub 2012-05-24.

44. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999 1999/01/01;6(1):1-55.

45. Kang T, Chen TT. Performance of the Generalized S-X^2 Item Fit Index for Polytomous IRT Models. Journal of Educational Measurement. 2008 2008/12/01;45(4):391-406.

46. Lameijer CM, van Bruggen SGJ, Haan EJA, Van Deurzen DFP, Van der Elst K, Stouten V, et al. Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® Upper Extremity V2.0 item bank in patients with upper extremity disorders. BMC Musculoskelet Disord [Internet]. 2020 2020/03//; 21(1):[170 p.]. Available from: http://europepmc.org/abstract/MED/32178644
https://doi.org/10.1186/s12891-020-3178-8
https://europepmc.org/articles/PMC7077019
https://europepmc.org/articles/PMC7077019?pdf=render.

47. McKinley RL, Mills CN. A Comparison of Several Goodness-of-Fit Statistics. Applied Psychological Measurement. 1985 1985/03/01;9(1):49-57.

48. Yen W, Fitzpatrick A. Item Response Theory. 2006.

49. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient. 2014;7(1):23-35.

50. Bortolotti SLV, Tezza R, de Andrade DF, Bornia AC, de Sousa Júnior AF. Relevance and advantages of using the item response theory. Quality & Quantity. 2013;47(4):2341-60.

51. Doostfatemeh M, Ayatollahi SMT, Jafari P. Assessing the effect of child's gender on their father–mother perception of the PedsQL™ 4.0 questionnaire: an iterative hybrid ordinal logistic regression/item response theory approach with Monte Carlo simulation. Health Qual Life Outcomes. 2020

2020/10/21;18(1):348.

52. Lee Sang E, Lee Pu R, Shin K-I. A composite estimator for stratified two stage cluster sampling. Communications for Statistical Applications and Methods. 2016 01/31;23(1):47-55. En.